

#22
8-21-02

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/015,078

DATE: 08/21/2002
 TIME: 14:04:28

Input Set : A:\235673-1.app
 Output Set: N:\CRF4\08212002\I015078.raw

SEQUENCE LISTING

4 (1) GENERAL INFORMATION:
 6 (i) APPLICANT: Suerbaum, Sebastian
 7 Labigne, Agnes
 9 (ii) TITLE OF INVENTION: Cloning and Characterization of the flbA
 10 Gene of H. Pylori, Production of Aflagellate Strains
 12 (iii) NUMBER OF SEQUENCES: 13
 14 (iv) CORRESPONDENCE ADDRESS:
 15 (A) ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
 16 Dunner
 17 (B) STREET: 1300 I Street, N.W.
 18 (C) CITY: Washington
 19 (D) STATE: D.C.
 20 (E) COUNTRY: USA
 21 (F) ZIP: 20005-3315
 23 (v) COMPUTER READABLE FORM:
 24 (A) MEDIUM TYPE: Floppy disk
 25 (B) COMPUTER: IBM PC compatible
 26 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 27 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30
 29 (vi) CURRENT APPLICATION DATA:
 C--> 30 (A) APPLICATION NUMBER: US/09/015,078
 C--> 31 (B) FILING DATE: 29-Jan-1998
 32 (C) CLASSIFICATION:
 34 (viii) ATTORNEY/AGENT INFORMATION:
 35 (A) NAME: Meyers, Kenneth J.
 36 (B) REGISTRATION NUMBER: 25,146
 37 (C) REFERENCE/DOCKET NUMBER: 02356.0073-01000
 39 (ix) TELECOMMUNICATION INFORMATION:
 40 (A) TELEPHONE: (202) 408-4000
 41 (B) TELEFAX: (202) 408-4400
 44 (2) INFORMATION FOR SEQ ID NO: 1:
 46 (i) SEQUENCE CHARACTERISTICS:
 47 (A) LENGTH: 19 base pairs
 48 (B) TYPE: nucleic acid
 49 (C) STRANDEDNESS: single
 50 (D) TOPOLOGY: linear
 52 (ii) MOLECULE TYPE: DNA (genomic)
 57 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:
 59 ATGCCNGGNA AAGCARATG
 61 (2) INFORMATION FOR SEQ ID NO: 2:
 63 (i) SEQUENCE CHARACTERISTICS:
 64 (A) LENGTH: 18 base pairs

ENTERED

19

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/015,078

DATE: 08/21/2002

TIME: 14:04:28

Input Set : A:\235673-1.app

Output Set: N:\CRF4\08212002\I015078.raw

```

65          (B) TYPE: nucleic acid
66          (C) STRANDEDNESS: single
67          (D) TOPOLOGY: linear
69      (ii) MOLECULE TYPE: DNA (genomic)
74      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:
76  RAAATTTCATN GCNCCRTC                                     18
78  (2) INFORMATION FOR SEQ ID NO: 3:
80      (i) SEQUENCE CHARACTERISTICS:
81          (A) LENGTH: 135 base pairs
82          (B) TYPE: nucleic acid
83          (C) STRANDEDNESS: single
84          (D) TOPOLOGY: linear
86      (ii) MOLECULE TYPE: DNA (genomic)
91      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
93  ATGCCAGGAA AGCAAATGGC GATTGATGCG GATTTAAATT CAGGGCTTAT TGATGATAAG      60
95  GAAGCTAAAA AACGGCGCGC CGCTCTAAGC CAAGAAGCGG ATTTTATG TGCGATGGAT      120
97  GCGCGCTCTA AATTT                                     135
99  (2) INFORMATION FOR SEQ ID NO: 4:
101      (i) SEQUENCE CHARACTERISTICS:
102          (A) LENGTH: 28 base pairs
103          (B) TYPE: nucleic acid
104          (C) STRANDEDNESS: single
105          (D) TOPOLOGY: linear
107      (ii) MOLECULE TYPE: DNA (genomic)
112      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
114  CGGGATCCGT GGTACTAAT GGTCTAC                                     28
116  (2) INFORMATION FOR SEQ ID NO: 5:
118      (i) SEQUENCE CHARACTERISTICS:
119          (A) LENGTH: 28 base pairs
120          (B) TYPE: nucleic acid
121          (C) STRANDEDNESS: single
122          (D) TOPOLOGY: linear
124      (ii) MOLECULE TYPE: DNA (genomic)
129      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:
131  CGGGATCCTC ATGGCCTCTT CAGAGACC                                     28
133  (2) INFORMATION FOR SEQ ID NO: 6:
135      (i) SEQUENCE CHARACTERISTICS:
136          (A) LENGTH: 2501 base pairs
137          (B) TYPE: nucleic acid
138          (C) STRANDEDNESS: single
139          (D) TOPOLOGY: linear
141      (ii) MOLECULE TYPE: DNA (genomic)
146      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
148  AGCTTTTTTG TGCCATACTT TTAAACTTTA TATTATAATA AGAGACAAAC ACACCTACCA      60
150  AAATTAAGGC ATTGATTTTA GATTATGGCA AACGAACGCT CCAAATTAGC TTTTAAAAAG      120
152  ACTTTCCTTG TCTTTAAACG CTTCTTGCAA TCCAAAGACT TAGCCCTTGT GGTCTTTGTG      180
154  ATAGCGATTT TAGCGATCAT TATCGTGCCG TTACCGCCTT TTGTGTTGGA TTTTCTACTC      240
156  ACGATTTCTA TCGCGCTATC GGTGTTGATT ATTTTAAATCG GGCTTTATAT TGACAAACCG      300
158  ACTGATTTTA GCGCTTTCCC CACTTTATTA CTCATTGTAA CCTTATACCG CTTGGCTTTA      360

```

RAW SEQUENCE LISTING

DATE: 08/21/2002

PATENT APPLICATION: US/09/015,078

TIME: 14:04:28

Input Set : A:\235673-1.app

Output Set: N:\CRF4\08212002\I015078.raw

```

160 AATGTCGCCA CCACTAGAAT GATTTTAAACC CAAGGCTATA AAGGGCCTAG CGCGGTGAGC 420
162 ATTATTATCA CGGCGTTTGG GGAATTTAGC GTGAGCGGGA ATTATGTGAT TGGGGCTATT 480
164 ATCTTTAGTA TTTTAGTGCT GGTGAATTTA TTAGTGGTTA CTAATGGTTC TACTAGGGTT 540
166 ACTGAAGTTA GGGCGCGATT TGCCCTAGAC GCTATGCCAG GAAAGCAAAT GGCGATTGAT 600
168 GCGGATTTAA ATTCAGGGCT TATTGATGAT AAGGAAGCTA AAAAACGGCG CGCCGCTCTA 660
170 AGCCAAGAAG CGGATTTTTA TGGTGCGATG GATGGCGCGT CTAAATTTGT CAAAGGCGAT 720
172 GCGATCGCTT CTATCATTAT CACGCTTATC AATATCATTG GGGGTTTTTT AGTGGGCGTG 780
174 TTCCAAAGGG ATATGAGCTT GAGCTTTAGT GCTAGCACTT TCACTATCTT AACCATTGGC 840
176 GATGGGCTTG TAGGGCAAAT CCCTGCCTTA ATCATTGCGA CACGGACCGG TATTGTCGCC 900
178 ACTCGCACCA CGCAAAACGA AGAAGAGGAC TTTGCTTCTA AGCTCATCAC ACAGCTCACC 960
180 AATAAAAGCA AAACCTTAGT GATTGTGGGG GCGATTTATT GCTTTTGCAC CATTCTGGA 1020
182 CTCCCTACCT TTTCTTTAGC GTTTGTAGGG GCTCTCTTTT TATTCATCGC ATGGCTGATT 1080
184 AGCAGGGAGG GAAAGGACGG GTTGCTCACT AAATTAGAAA ATTATTTGAG TCAAAAATTC 1140
186 GGCTTGGATT TGAGCGAAAA ACCCCACAGC TCCAAAATCA AACCCACGC CCCACCACA 1200
188 AGGGCTAAAA CCCAAGAAGA GATTAAGAAG GAAGAAGAGC AAGCCATTGA TGAAGTGTTA 1260
190 AAAATTGAAT TTTTAGAATT GGCTTTAGGC TATCAGCTCT ACAGCTTAGC GGACATGAAA 1320
192 CAAGGGGGCG ATTTGTTAGA AAGGATTAGG GGTATTAGAA AAAAGATAGC GAGCGATTAT 1380
194 GGTTTTTTGA TGCCTCAAAT TAGGATTAGG GATAATTTAC AACTCCCCC AACGCATTAT 1440
196 GAAATCAAGC TTAAGGGCAT TGTGATTGGT GAAGGCATGG TGATGCCGGA TAAGTTTTTA 1500
198 GCCATGAATA CCGGTTTTGT GAATAAAGAA ATTGAAAGCA TTCTACTAA AGAGCCGGCT 1560
200 TTTGGAATGG ACGCTTTATG GATTGAAACT AAAAAATAAG AAGAAGCCAT CATTCAAGGC 1620
202 TATACCATTA TTGATCCAAG CACCGTTATT GCGACGCACA CCAGCGAATT AGTGAAAAAA 1680
204 TACGCTGAAG ATTTTATCAC TAAAGATGAA GTGAAATCCC TTTTAGAGCG CTTGGCCAAA 1740
206 GACTATCCTA CGATTGTAGA AGAGAGTAAA AAAATCCCCA CCGGTGCGAT CCGATCAGTC 1800
208 TTGCAAGCCT TGTTGCATGA AAAAATCCCC ATTAAGACA TGCTCACTAT TTTAGAAACG 1860
210 ATTACCGATA TTGCGCCATT AGTTCAAAAC GATGTGAATA TCTTAACCGA ACAAGTGAGG 1920
212 GCGAGGCTTT CTAGGGTGAT CACTAACGCT TTTAAATCTG AAGACGGGCG TTTGAAATTT 1980
214 TTAACCTTTT CTACCGATAG CGAACAAATT TTGCTTAATA AATGCGGAGA AAATGGCACT 2040
216 TCTAAGAGCC TACTACTCAA TGTGGGCGAA TTGCAAAAAC TCATTGAAGC GGTCTCTGAA 2100
218 GAGGCCATGA AAGTCTTGCA AAAAGGGATC GCTCCGGTGA TTTTGATCGT AGAGCCTAAT 2160
220 TTAAGAAAAG CCCTTCTTAA TCAAATGGAG CAGGCTAGGA TTGATGTAAT CGTGCTAAGC 2220
222 CATGCTGAAT TAGATCCTAA CTCTAATTTT GAAGCCTTAG GCACGATCCA TATTAACCTT 2280
224 TAAGGGATAA ATAATTGATA AAAAAGGAGA ATGATGCAAG TTTATCACCT TTCACACATT 2340
226 GATTTAGACG GCTATGCATG CCAGCTTGTT TCAAAACAAT TTTTAAAAA TATCCAATGC 2400
228 TATAACGCTA ATTACGGGCG TGAAGTCTCA GCGAGAATTT ATGAGATTTT AAACGCGATC 2460
230 GCTCAATCTA AAGAGAGTGA ATTCCTTATT TTGATTAGCG A 2501

```

232 (2) INFORMATION FOR SEQ ID NO: 7:

234 (i) SEQUENCE CHARACTERISTICS:

235 (A) LENGTH: 732 amino acids

236 (B) TYPE: amino acid

237 (C) STRANDEDNESS: single

238 (D) TOPOLOGY: linear

240 (ii) MOLECULE TYPE: peptide

245 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 7:

```

247 Met Ala Asn Glu Arg Ser Lys Leu Ala Phe Lys Lys Thr Phe Pro Val
248 1 5 10 15
250 Phe Lys Arg Phe Leu Gln Ser Lys Asp Leu Ala Leu Val Val Phe Val
251 20 25 30
253 Ile Ala Ile Leu Ala Ile Ile Ile Val Pro Leu Pro Pro Phe Val Leu

```

RAW SEQUENCE LISTING
 PATENT APPLICATION: US/09/015,078

DATE: 08/21/2002
 TIME: 14:04:28

Input Set : A:\235673-1.app
 Output Set: N:\CRF4\08212002\I015078.raw

254		35		40		45										
256	Asp	Phe	Leu	Leu	Thr	Ile	Ser	Ile	Ala	Leu	Ser	Val	Leu	Ile	Ile	Leu
257		50					55					60				
259	Ile	Gly	Leu	Tyr	Ile	Asp	Lys	Pro	Thr	Asp	Phe	Ser	Ala	Phe	Pro	Thr
260	65					70					75				80	
262	Leu	Leu	Leu	Ile	Val	Thr	Leu	Tyr	Arg	Leu	Ala	Leu	Asn	Val	Ala	Thr
263					85					90				95		
265	Thr	Arg	Met	Ile	Leu	Thr	Gln	Gly	Tyr	Lys	Gly	Pro	Ser	Ala	Val	Ser
266				100					105					110		
268	Ile	Ile	Ile	Thr	Ala	Phe	Gly	Glu	Phe	Ser	Val	Ser	Gly	Asn	Tyr	Val
269			115					120					125			
271	Ile	Gly	Ala	Ile	Ile	Phe	Ser	Ile	Leu	Val	Leu	Val	Asn	Leu	Leu	Val
272		130					135					140				
274	Val	Thr	Asn	Gly	Ser	Thr	Arg	Val	Thr	Glu	Val	Arg	Ala	Arg	Phe	Ala
275	145					150					155				160	
277	Leu	Asp	Ala	Met	Pro	Gly	Lys	Gln	Met	Ala	Ile	Asp	Ala	Asp	Leu	Asn
278					165					170				175		
280	Ser	Gly	Leu	Ile	Asp	Asp	Lys	Glu	Ala	Lys	Lys	Arg	Arg	Ala	Ala	Leu
281				180					185					190		
283	Ser	Gln	Glu	Ala	Asp	Phe	Tyr	Gly	Ala	Met	Asp	Gly	Ala	Ser	Lys	Phe
284			195					200				205				
286	Val	Lys	Gly	Asp	Ala	Ile	Ala	Ser	Ile	Ile	Ile	Thr	Leu	Ile	Asn	Ile
287		210					215					220				
289	Ile	Gly	Gly	Phe	Leu	Val	Gly	Val	Phe	Gln	Arg	Asp	Met	Ser	Leu	Ser
290	225					230					235				240	
292	Phe	Ser	Ala	Ser	Thr	Phe	Thr	Ile	Leu	Thr	Ile	Gly	Ala	Gly	Leu	Val
293					245					250				255		
295	Gly	Gln	Ile	Pro	Ala	Leu	Ile	Ile	Ala	Thr	Arg	Thr	Gly	Ile	Val	Ala
296			260					265					270			
298	Thr	Arg	Thr	Thr	Gln	Asn	Glu	Glu	Glu	Asp	Phe	Ala	Ser	Lys	Leu	Ile
299		275					280						285			
301	Thr	Gln	Leu	Thr	Asn	Lys	Ser	Lys	Thr	Leu	Val	Ile	Val	Gly	Ala	Ile
302		290				295						300				
304	Tyr	Cys	Phe	Cys	Thr	Ile	Pro	Gly	Leu	Pro	Thr	Phe	Ser	Leu	Ala	Phe
305	305					310					315				320	
307	Val	Gly	Ala	Leu	Phe	Leu	Phe	Ile	Ala	Trp	Leu	Ile	Ser	Arg	Glu	Gly
308					325					330				335		
310	Lys	Asp	Gly	Leu	Leu	Thr	Lys	Leu	Glu	Asn	Tyr	Leu	Ser	Gln	Lys	Phe
311			340					345					350			
313	Gly	Leu	Asp	Leu	Ser	Glu	Lys	Pro	His	Ser	Ser	Lys	Ile	Lys	Pro	His
314		355					360					365				
316	Ala	Pro	Thr	Thr	Arg	Ala	Lys	Thr	Gln	Glu	Glu	Ile	Lys	Arg	Glu	Glu
317		370					375					380				
319	Glu	Gln	Ala	Ile	Asp	Glu	Val	Leu	Lys	Ile	Glu	Phe	Leu	Glu	Leu	Ala
320	385					390					395				400	
322	Leu	Gly	Thr	Gln	Leu	Tyr	Ser	Leu	Ala	Asp	Met	Lys	Gln	Gly	Gly	Asp
323				405					410					415		
325	Leu	Leu	Glu	Arg	Ile	Arg	Gly	Ile	Arg	Lys	Lys	Ile	Ala	Ser	Asp	Tyr
326			420					425					430			

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/015,078

DATE: 08/21/2002

TIME: 14:04:28

Input Set : A:\235673-1.app

Output Set: N:\CRF4\08212002\I015078.raw

```

328 Gly Phe Leu Met Pro Gln Ile Arg Ile Arg Asp Asn Leu Gln Leu Pro
329      435      440      445
331 Pro Thr His Tyr Glu Ile Lys Leu Lys Gly Ile Val Ile Gly Glu Gly
332      450      455      460
334 Met Val Met Pro Asp Lys Phe Leu Ala Met Asn Thr Gly Phe Val Asn
335      465      470      475      480
337 Lys Glu Ile Glu Gly Ile Pro Thr Lys Glu Pro Ala Phe Gly Met Asp
338      485      490      495
340 Ala Leu Trp Ile Glu Thr Lys Asn Lys Glu Glu Ala Ile Ile Gln Gly
341      500      505      510
343 Tyr Thr Ile Ile Asp Pro Ser Thr Val Ile Ala Thr His Thr Ser Glu
344      515      520      525
346 Leu Val Lys Lys Tyr Ala Glu Asp Phe Ile Thr Lys Asp Glu Val Lys
347      530      535      540
349 Ser Leu Leu Glu Arg Leu Ala Lys Asp Tyr Pro Thr Ile Val Glu Glu
350      545      550      555      560
352 Ser Lys Lys Ile Pro Thr Gly Ala Ile Arg Ser Val Leu Gln Ala Leu
353      565      570      575
355 Leu His Glu Lys Ile Pro Ile Lys Asp Met Leu Thr Ile Leu Glu Thr
356      580      585      590
358 Ile Thr Asp Ile Ala Pro Leu Val Gln Asn Asp Val Asn Ile Leu Thr
359      595      600      605
361 Glu Gln Val Arg Ala Arg Leu Ser Arg Val Ile Thr Asn Ala Phe Lys
362      610      615      620
364 Ser Glu Asp Gly Arg Leu Lys Phe Leu Thr Phe Ser Thr Asp Ser Glu
365      625      630      635      640
367 Gln Phe Leu Leu Asn Lys Leu Arg Glu Asn Gly Thr Ser Lys Ser Leu
368      645      650      655
370 Leu Leu Asn Val Gly Glu Leu Gln Lys Leu Ile Glu Ala Val Ser Glu
371      660      665      670
373 Glu Ala Met Lys Val Leu Gln Lys Gly Ile Ala Pro Val Ile Leu Ile
374      675      680      685
376 Val Glu Pro Asn Leu Arg Lys Ala Leu Ser Asn Gln Met Glu Gln Ala
377      690      695      700
379 Arg Ile Asp Val Ile Val Leu Ser His Ala Glu Leu Asp Pro Asn Ser
380      705      710      715      720
382 Asn Phe Glu Ala Leu Gly Thr Ile His Ile Asn Phe
383      725      730
385 (2) INFORMATION FOR SEQ ID NO: 8:
387 (i) SEQUENCE CHARACTERISTICS:
388 (A) LENGTH: 732 amino acids
389 (B) TYPE: amino acid
390 (C) STRANDEDNESS: single
391 (D) TOPOLOGY: linear
393 (ii) MOLECULE TYPE: peptide
398 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 8:
400 Met Ala Asn Glu Arg Ser Lys Leu Ala Phe Lys Lys Thr Phe Pro Val
401 1 5 10 15
403 Phe Lys Arg Phe Leu Gln Ser Lys Asp Leu Ala Leu Val Val Phe Val

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/015,078

DATE: 08/21/2002

TIME: 14:04:29

Input Set : A:\235673-1.app

Output Set: N:\CRF4\08212002\I015078.raw

L:30 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:31 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]